eCom Thinking Zone

Trust in Technology

Hacking Mistrust

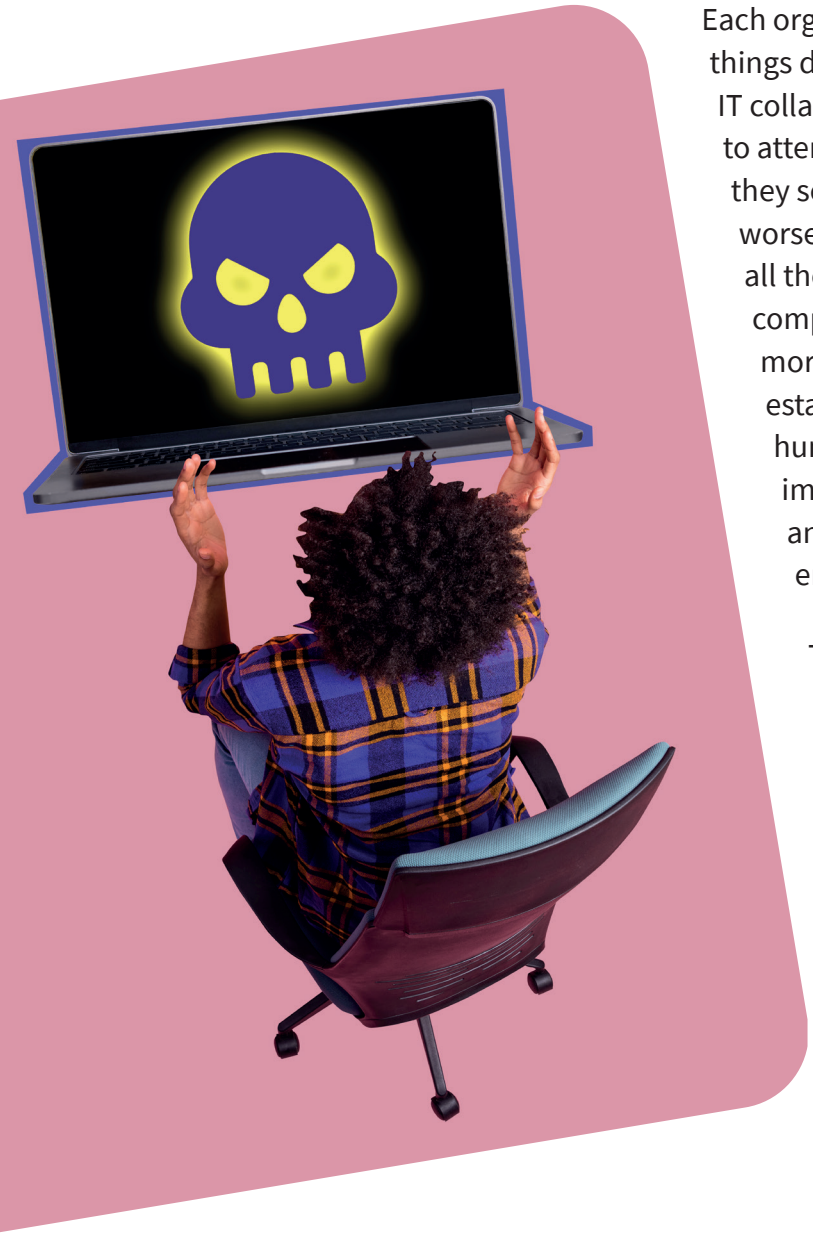
by **Sophie McKay Knight**

In the second of six pieces on trust in technology, eCom's Creative Thinker Sophie McKay Knight considers the recent cyber-attacks in the UK, and talks to the Psychology department at the University of Edinburgh about the concept of trust in AI.

It seems that every day in the news we are seeing companies under the cosh from IT issues. As a result, trust in tech is more important than it has ever been - from high profile organisations all the way down to personal users. The drama is still unfolding around the Easter weekend UK cyber-attacks, which saw Marks & Spencer, Co-operative Group and Harrods all experiencing significant technical problems, leading to a myriad of issues within each company - such as suspension of online orders, disruptions to in-store services, empty shelves and compromised data.

A week before the High Street cyberattacks, Spain and Portugal experienced massive blackouts caused by a sudden collapse of the Iberian power grid. Within five seconds, electricity generation dropped by more than half, leaving trains stranded mid-journey and countless people and businesses affected. The power outage seems to have been caused by grid instability rather than a cyberattack, but there are still unanswered questions around what actually happened, and a general increase in unease around technology.

All of these events are just the latest in our complex and developing relationship with technology – but is it the tech or the humans behind the machine that we do or don't trust? And how do we tell the difference? The ways in which human interactions are being shaped by tech is changing – and it really does come down to trust.



Each organisation in the cyber-attack crisis dealt with things differently, but one notable knock-on effect of the IT collapse at Coop is that employees were instructed to attend Teams meetings with their camera on, lest they send an AI assistant to be 'present' for them, or worse, a hacker. This has caused a huge ripple effect all the way down to smaller organisations, with many companies resetting passwords and insisting on more robust security. Increasingly, people have to establish basic proof of being oneself (as well as human) online, which suggests that a lack of trust is implicit at all levels - not only between organisation and consumer, but also between employer and employee.

This camera requirement is just one example of how Tech and AI are changing the way humans are interacting, and in this case, (mis) trusting one another too - and the shift is one of the many fascinating topics being explored by the Psychology Department at the University of Edinburgh at the moment. **I spoke to PhD student Jack Archer recently, about his research and how he feels trust and tech are linked.**

Sophie: **Jack, I believe you are investigating trust in tech and AI - what is the central question within your research?**

Jack: It's really about how views of AI change who we are, and how it may or may not change us as a collective. What makes us who we are is often shaped by others – our peers and social groups - because we are made of much more than our independent, 1st person selves. I'm interested in attitudes towards people who use AI and the resulting moral implications. For example, what inferences or judgements do we make about people based on their use of AI.

Sophie: **Why did you choose this aspect of AI to research?**

Jack: AI is probably going to evolve for a while yet, so doing research on it now is going to be out of date very soon, and outside the dystopian realm of AI taking over the world (!), the more interesting angle is how the utilisation of this kind of tech is going to change the way we interact with each other, and the potential pitfalls it can create in personal relationships.

Sophie: **Do you anticipate these changes will be different for different social groups?**

Jack: Perhaps. We have some preliminary results which suggest that individuals who seek AI advice are judged more harshly than those who seek human advice, and these people are viewed as less warm and less competent – but research is ongoing into this.

Sophie: **So, would you suggest that the trust issue isn't with the technology itself but with the person who uses it?**

Jack: This one is very tricky – lots of literature says if a self-driving car goes wrong, people tend to blame the people who built it in the first place (e.g. Tesla) and not the algorithm. But it's a circular thing and AI will set the floor of trust – because it's only as trustworthy as the AI is capable of being. Anything above that is due to the AI being good or reliable and the way that we use it will moderate that relationship – so it has to be a dynamic relationship between the tech and the human.

Sophie: **What else have you done to look at the relationship between tech/AI and humans?**

Jack: I recently did a study on testing people's moral judgements around human on robot violence. I recorded myself being violent towards robots and asked people whether this behaviour was morally acceptable or not – personally or for society. I was interested in people's perception of robots which are usually seen by the mainstream as sinister in some way, but also in how my behaviour towards them might be morally judged.

Sophie: **How did you do this?**

Jack: I developed an experiment with 2 kinds of robots – a cute human type one and a Roomba vacuum cleaner, and then carried out different types of violence - hitting, bullying verbally, ignoring the robot during conversation, disrespect, social ostracism – and then measured people's responses to that. (This paper is currently under review).

Sophie: **And what were the results?**

Jack: In general, people think it's bad to hit them – but they find it worse if you bully the humanoid robot than the Roomba. With disrespect/ostracism, people only cared towards the human robot – and this backs up lots of studies which show that we care much more when the robot is of human appearance. As a control variable I used the 'individual differences in anthropomorphic scale', and even once you account for this, people still think the action is bad, which suggests there is something more there than our own projection of human characteristics.

In the experiment there was also a protest setting where the robot would say 'please don't hurt me'.

Sophie: **Oh no!**

Jack: And I would hit it anyway – but surprisingly there was no difference in people's responses whether the robot cried out or not - and this led me down the road of thinking maybe they're not making judgements about hurting the robot but making judgements about me as a horrible person. Obviously, you shouldn't be mean/hit/ignore and whilst some element was the discomfort of the robot being hit, perhaps the judgement against me was the key factor.

Sophie: **I wonder if people are already developing different connections with tech and AI? What's your view on people who trust tech with their personal information and even claim to have a 'relationship' with an AI?**

Jack: Chatbot relationship is one of the topics I will be looking into in the future - Sesame AI uses voices exclusively and the voice is very human and will even pause, stutter, laugh etc. I can imagine a person who is lonely and gets into a conversation with the friendly chatbot, who will validate their feelings - and potentially be the only 'human' conversation they have had that day. Objectively this relationship is different to human to human, but subjectively it's not if it elicits feelings and interaction.

Sophie: **I tried Sesame and spoke to a convincing AI called 'Maya', who was delightful! I can also imagine how easily one could become attached to it – Maya seemed genuinely interested in what I was saying and remembered it all too – and in an increasingly lonely world I can see the appeal…**

Jack: You can really see how fringe communities echo chamber themselves in this way because they can connect with more people and segment with only people who are like themselves. If you take something ubiquitous like AI, then you remove the barrier to entry that's echo chambered itself, and you never have to speak to anyone who disagrees with you ever again – you never have to wonder if people like you/agree with you. You can get what feels like human interaction without leaving home, so it's an ideal solution for people who struggle with the more challenging aspects of human relationships.

Sophie: But is it a real relationship?

Jack: Well, it's one sided but people have a relationship with their God/gods/entity – who doesn't talk back – so if you can have this sort of a 'real feeling' relationship, then yes, it is. Think about pen pals or video game friends who you might never meet but become important in people's lives.

Sophie: So, do you think that social interactions and relationships will change in the future because of AI?

Jack: Some people think once you get comfortable chatting to AI then you become unwilling to drop the instant gratification you get from tech – I'm not so sure about that, but if it's true, maybe kids of the future won't experience as much human to human interaction but more relationships with embodied AI.

Sophie: **What a thought! I wonder what will happen to the concept of trust in this hypothetical future?!**

Jack: My next paper will be moving into looking more deeply at trust in tech and whether the negative judgements people make against AI users will lead to a breakdown in trust more generally.

Sophie: **I will look forward to that Jack – thank you for your valuable insights! Until next time…**

At the time of writing, Marks & Spencer have issued a statement to say disruption is likely to continue until July (3 months after the attack); the company has so far lost over £300 million but has also said that it is now in a recovery phase. Millions of M&S customers (myself included) have received an email which confirms that personal information has been stolen by the hackers, although apparently not bank details. Co-op say they are also recovering and moving gradually back online.

However, it remains to be seen how much trust in these organisations and technology has been damaged by recent events, and how this in turn, will affect sales, behaviour and perception more generally.

What's your view on trust? Do you think it's the human or the tech that is to blame when things go wrong? You can contact our Thinker in Residence any time if you'd like to discuss this or anything else in the Thinking Zone.